

# From Containers to Content to Context: the Changing Role of Libraries in eScience and eScholarship

Prof. Dr. Stefan Gradmann  
Humboldt-Universität zu Berlin / School of Library and Information Science  
Präsident der Deutschen Gesellschaft für Informationswissenschaft und  
Informationspraxis (DGI)  
[stefan.gradmann@ibi.hu-berlin.de](mailto:stefan.gradmann@ibi.hu-berlin.de)



# Overview



## **The Poet, the Library and the Scriptorium**

How libraries and content were once closely connected

## **The Gutenberg Parenthesis opens ...**

Dissociation of container and content in the print paradigm

## **... and closes again**

The end of the print paradigm

## **Documents and Data in eScience and eScholarship**

Context is king!

## **An Opportunity for Libraries ...**

... and what they need to do to be up to it



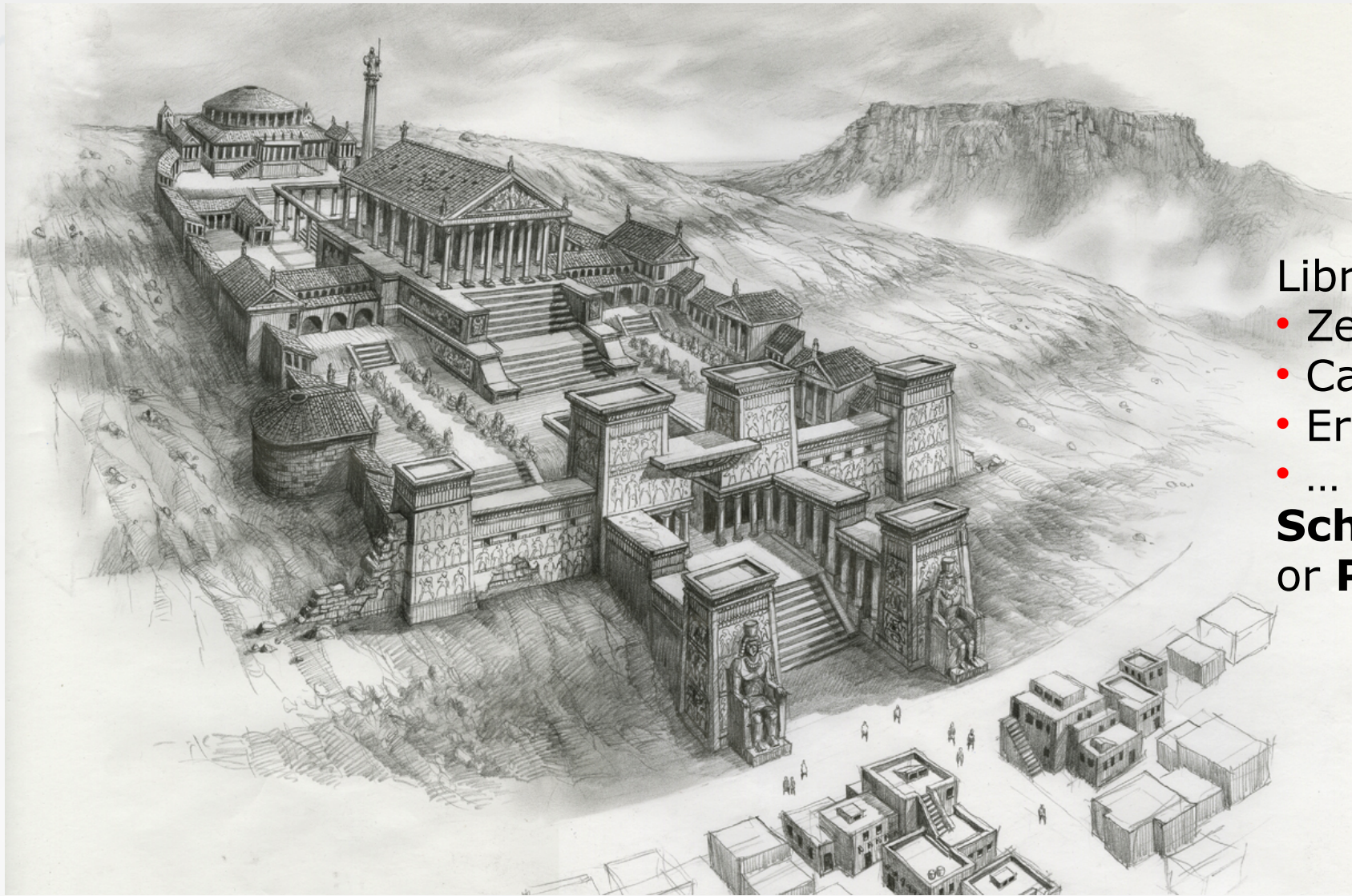
# The Poet, the Library and the Scriptorium

How libraries and content were connected before the Gutenberg Parenthesis





# Long before the Parenthesis: Alexandria



Librarians:

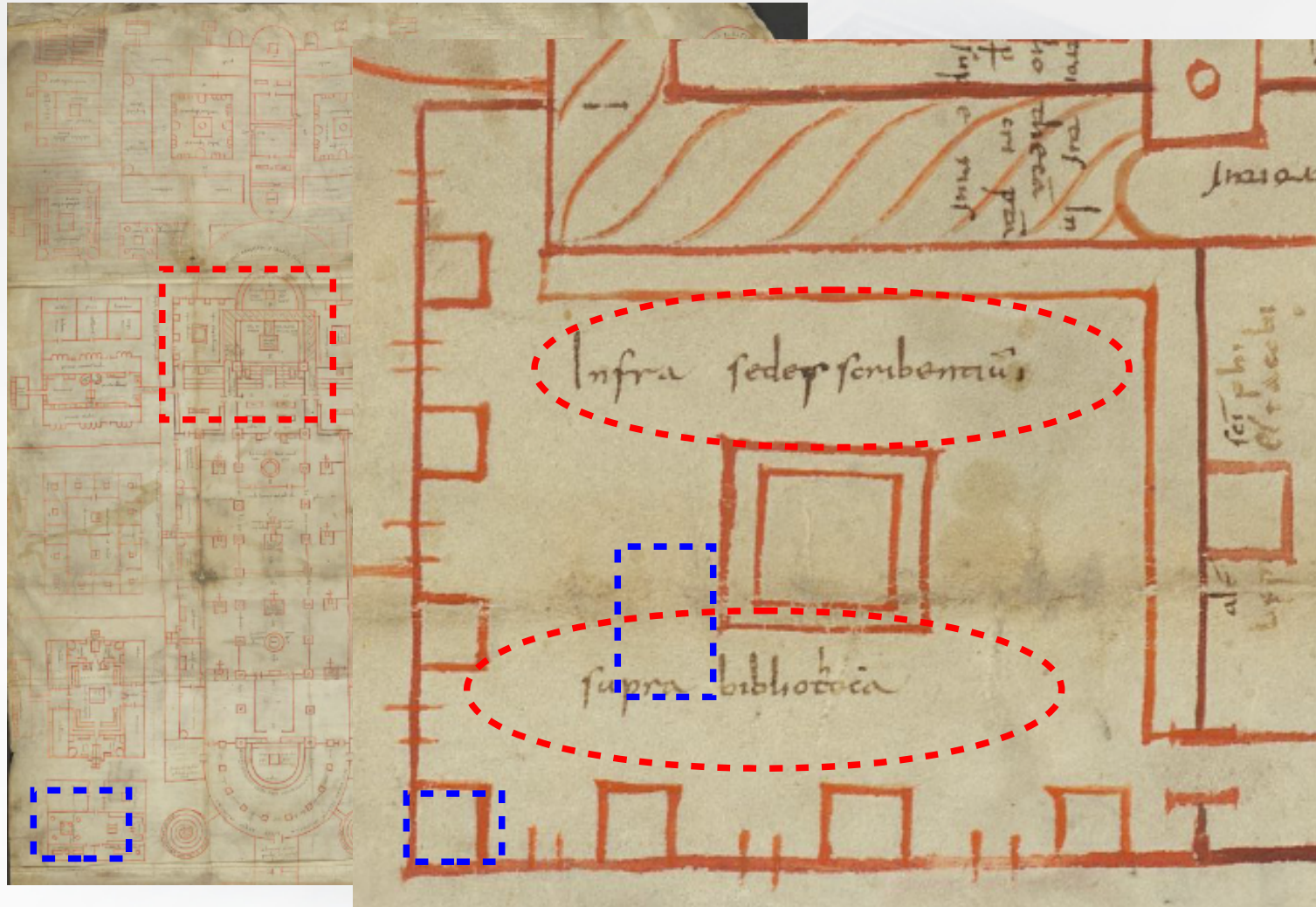
- Zenodotus
- Callimachus
- Erathosthenes
- ...

**Scholars** and /  
or **Poets**





# Before the Parenthesis: St. Gall



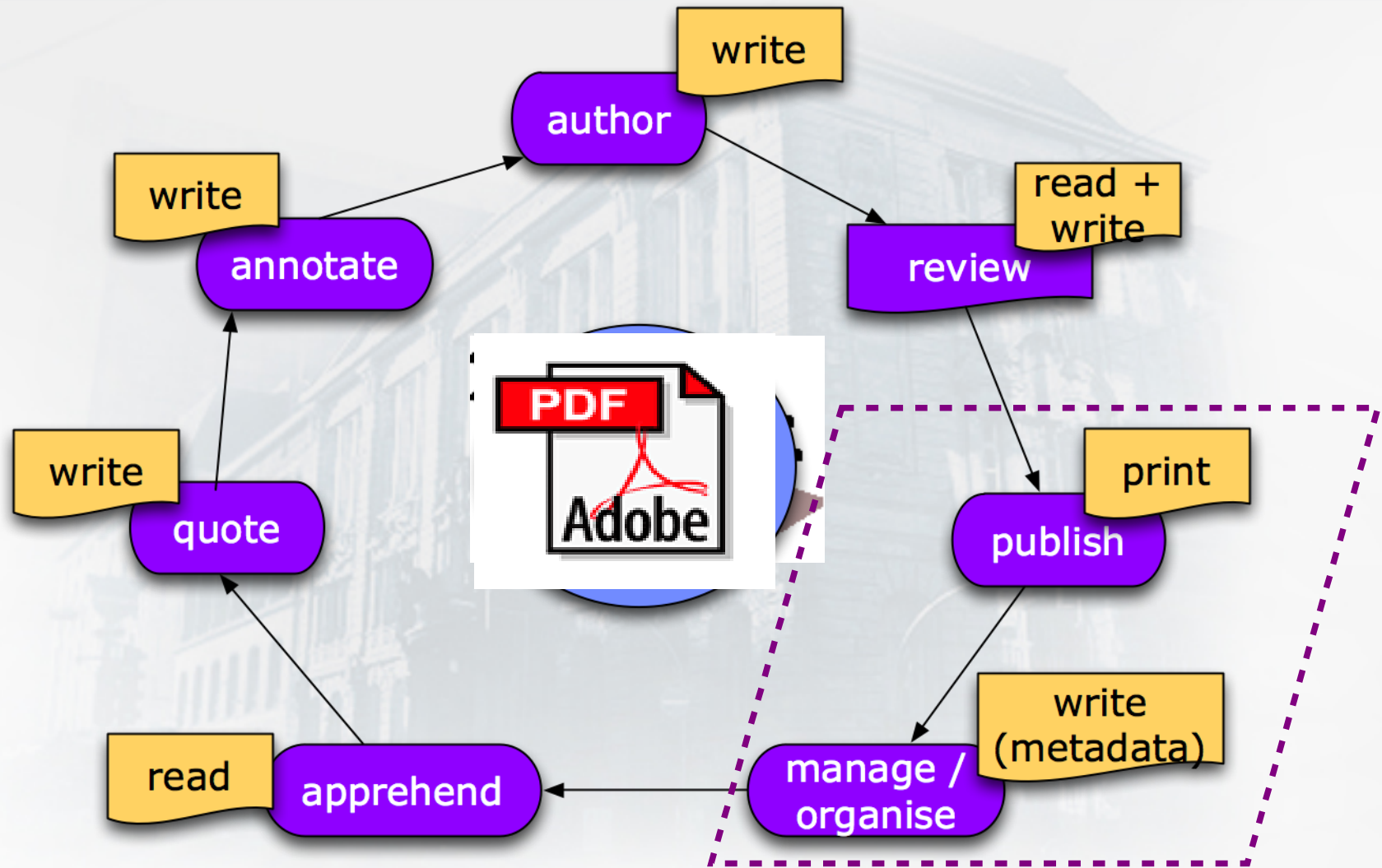
# The Gutenberg Parenthesis Opens ...

## Dissociation of container and content in the print paradigm

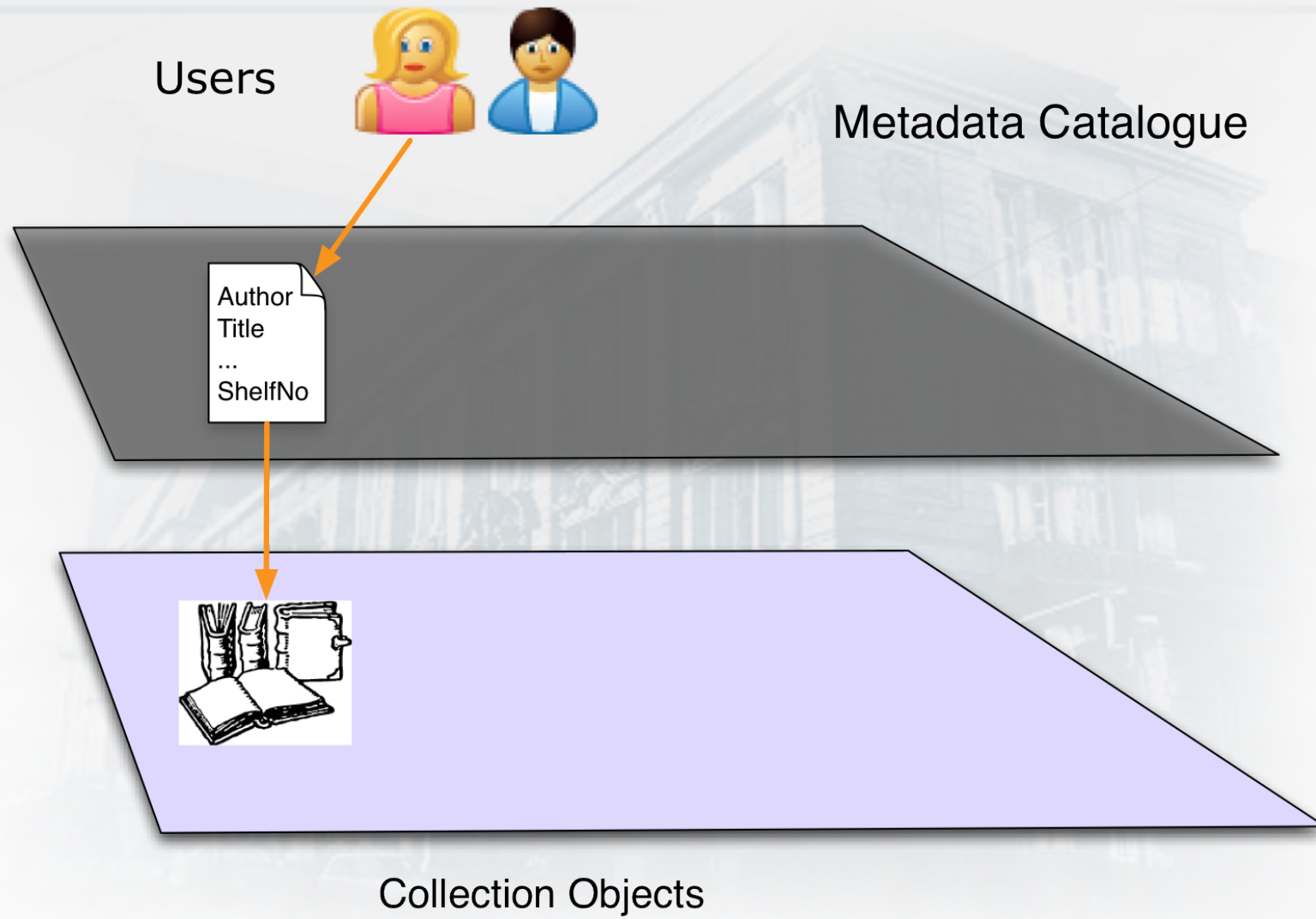




# Dissociation of Roles in the Gutenberg galaxy

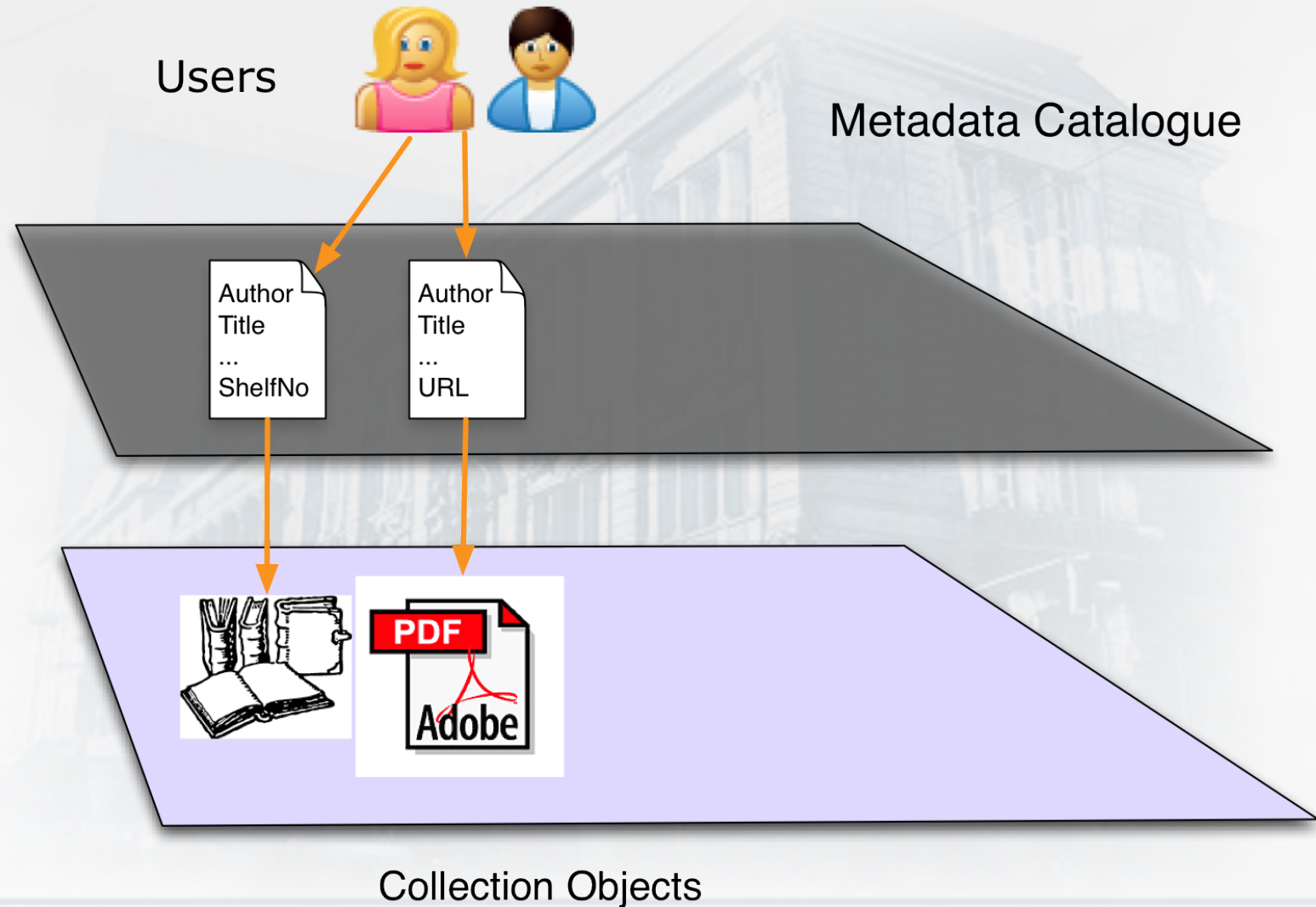


# Catalogue Based Library Functional Axioms (1)

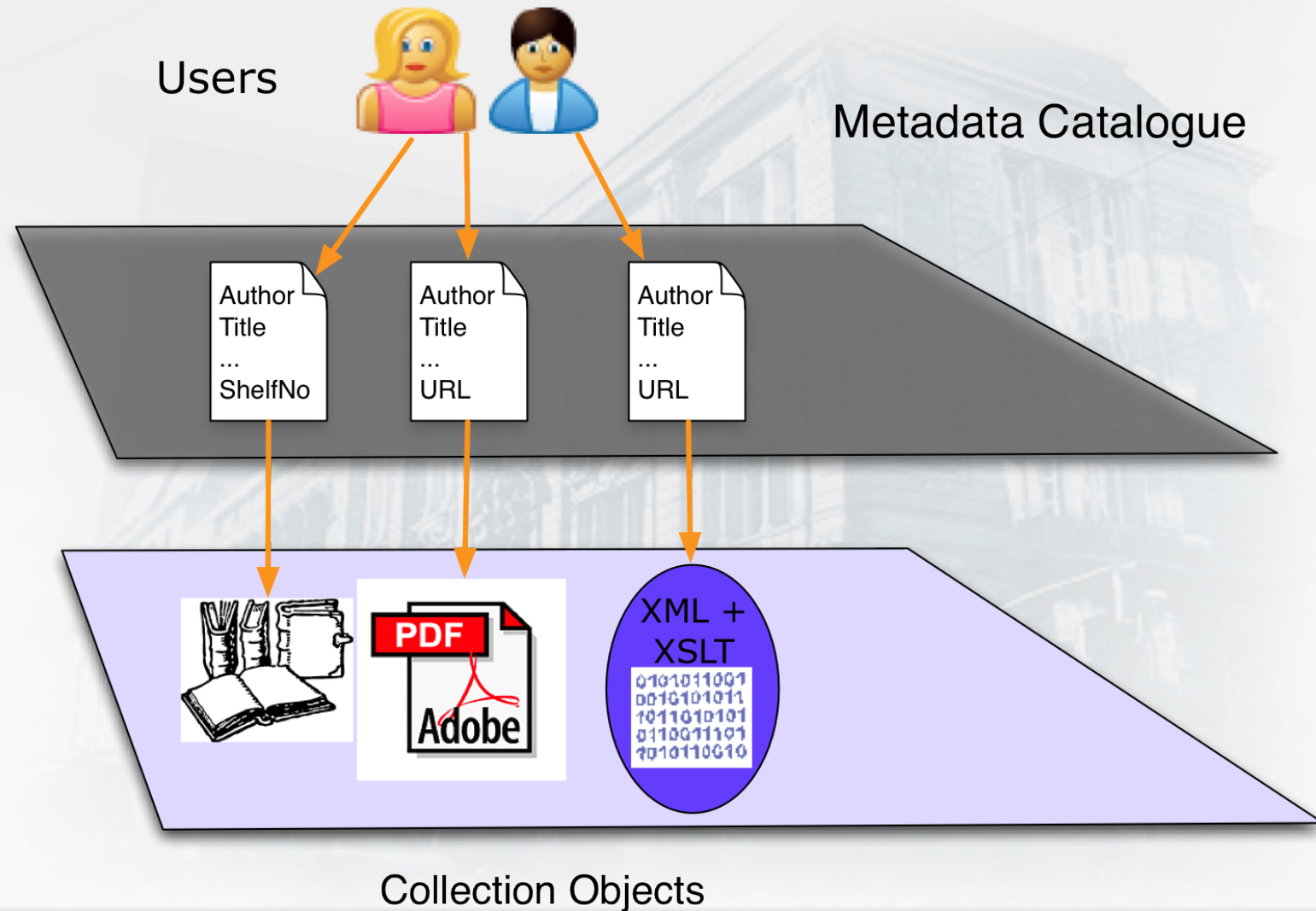




# Catalogue Based Library Functional Axioms (2)

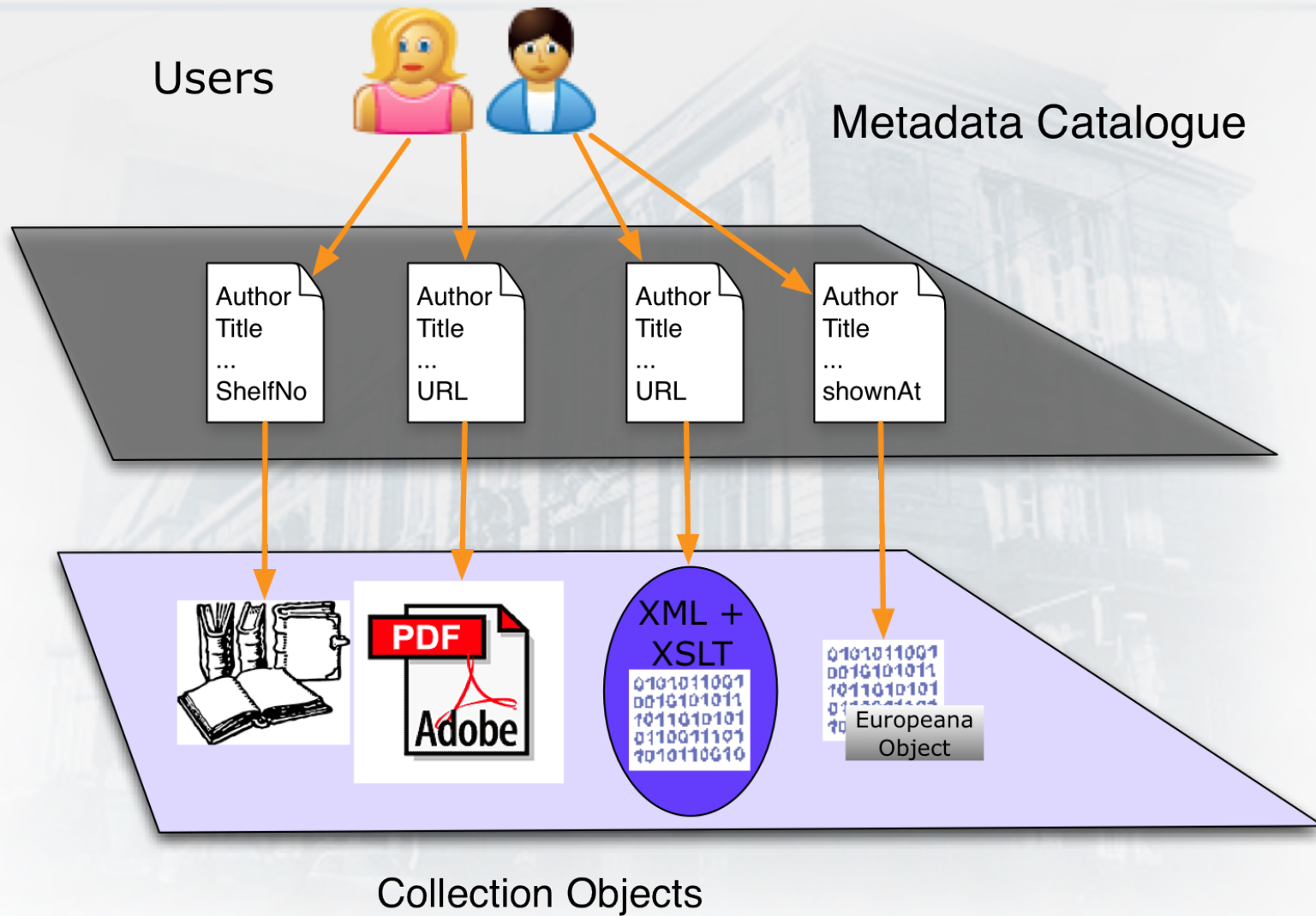


# Catalogue Based Library Functional Axioms (3)





# Catalogue Based Library Functional Axioms (4)



# Library Functional Principles (5)



- **Mediating access** to information objects via **catalogues**
- **Mediating links** as pointers from metadata to objects
- Objects are part of a library **collection**
  - An object to be used within a library typically is part of this library's collection
- Internal processing logic: focus on
  - objects as information **containers**,
  - not so much on the **content** of these containers
  - and accordingly **cataloguing** is focussed on **container attributes**
- Functional macro-primitives are **ingestion**, **storage**, **description** and **retrieval** of information containers



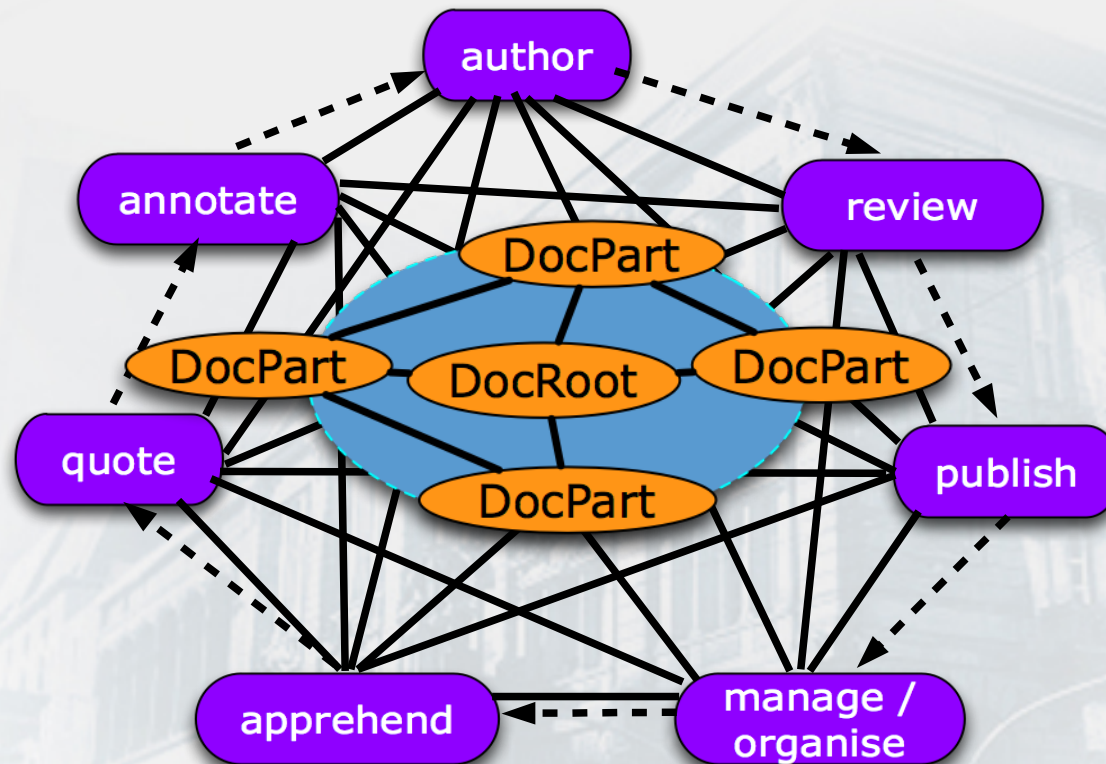
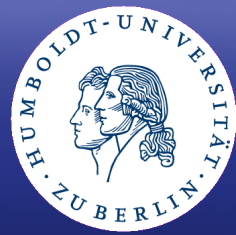


# **... and closes again**

## The end of the print paradigm



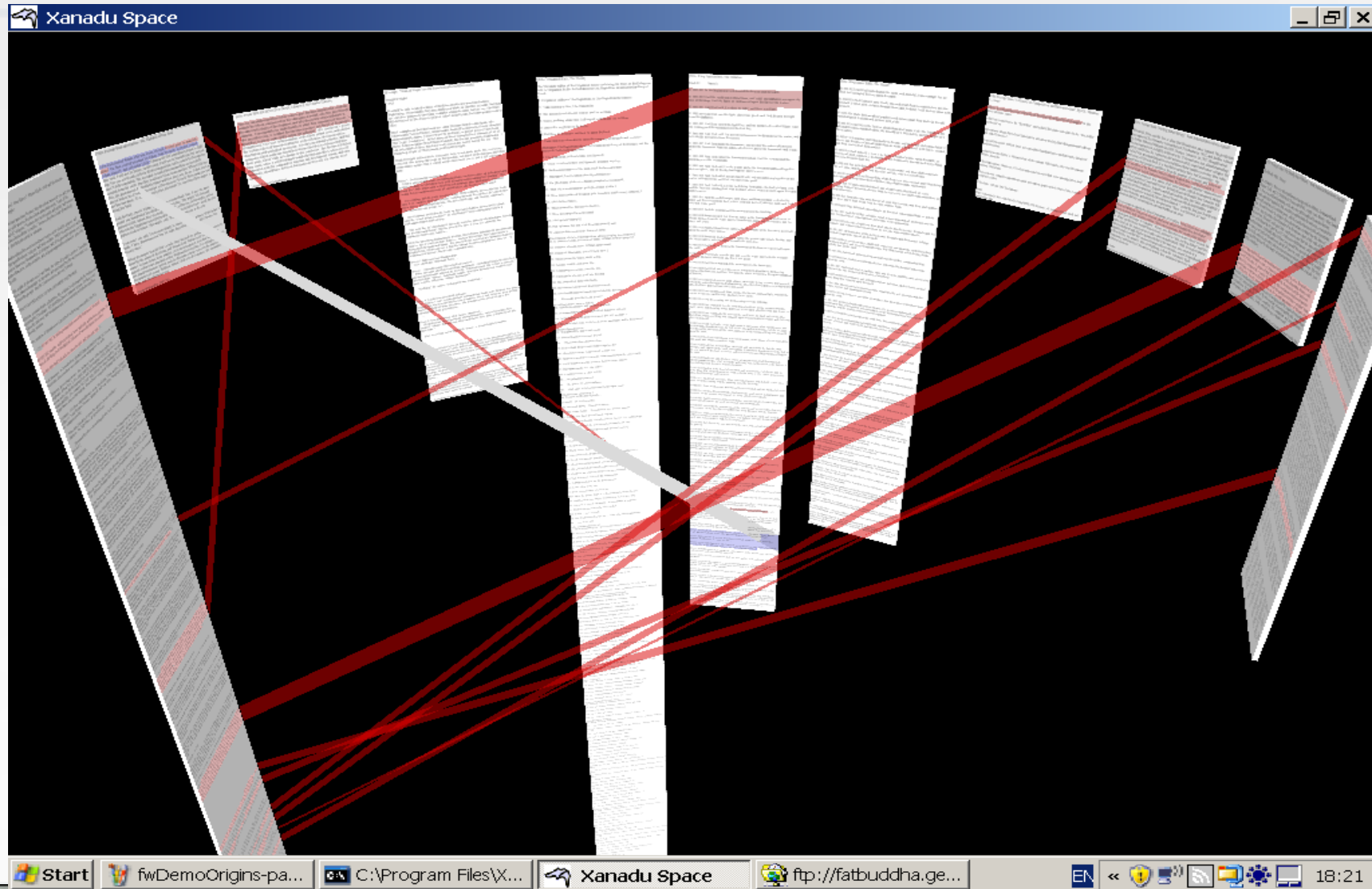
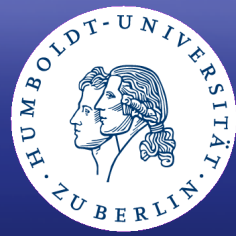
# Web Based Scholarly Working Continuum ... ... a triple paradigm shift: Beyond Documents



- Decreasing functional determination by traditional cultural techniques
- Disintegration of the linear / circular functional paradigma
- Erosion of the monolithic document notion in hypertext paradigms

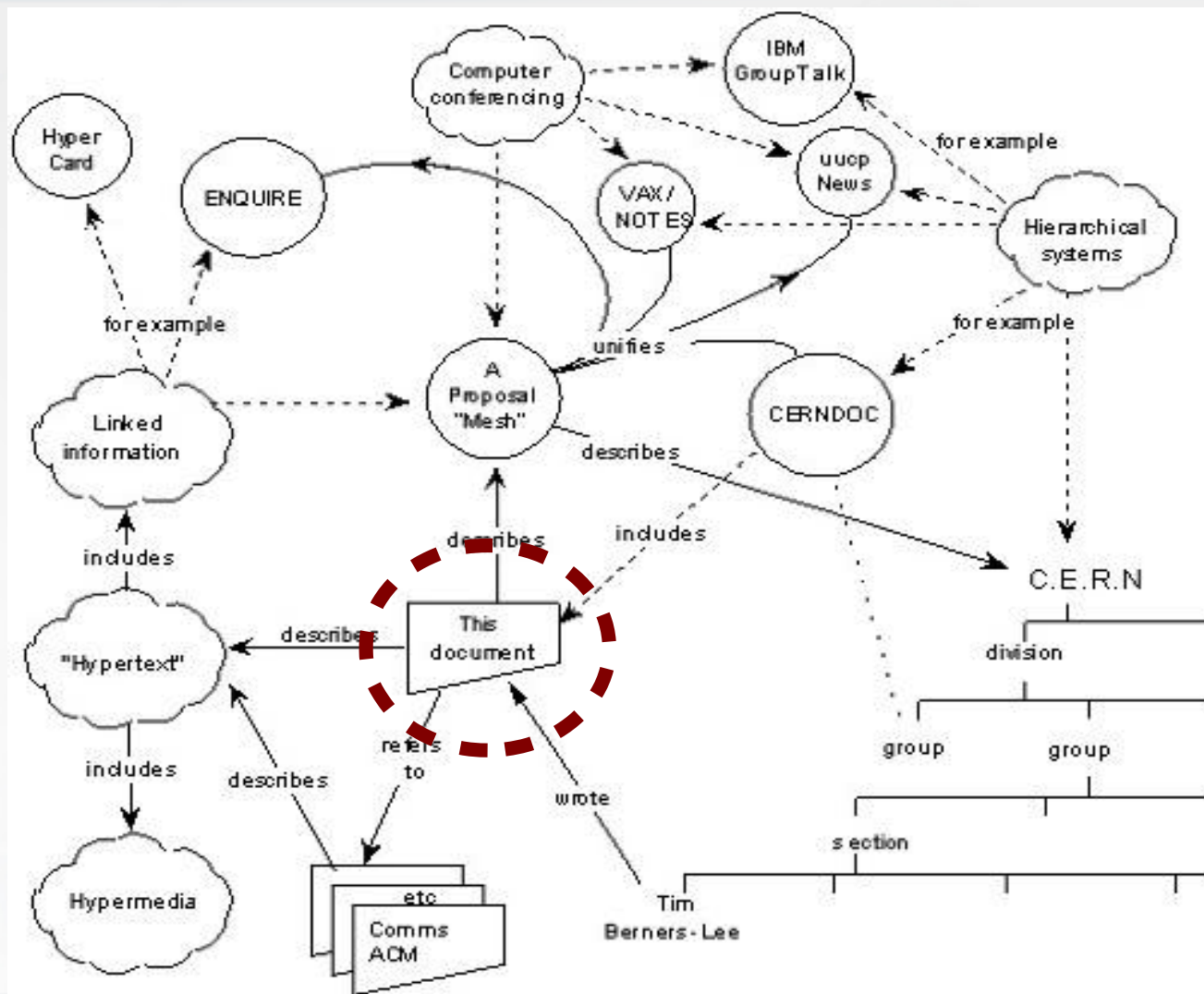


# Ted Nelson's Xanadu: radicalised Hypertext ...



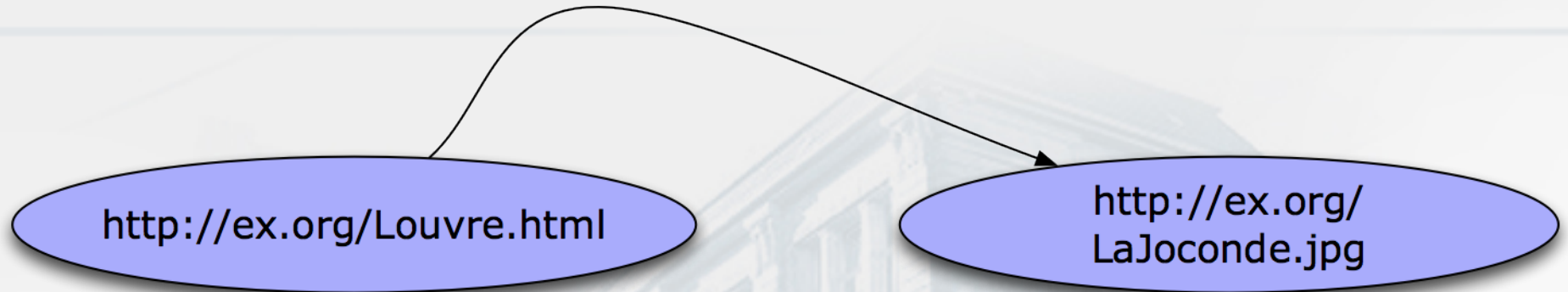


# The Web of Documents



Information Management:  
A Proposal  
(TBL, 1989)

# Resources and Links in the Document Web

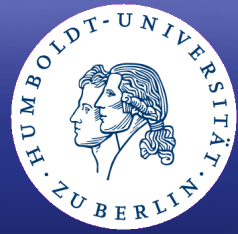


```
<a href="http://ex.org/LaJoconde.jpg">b</a>
```

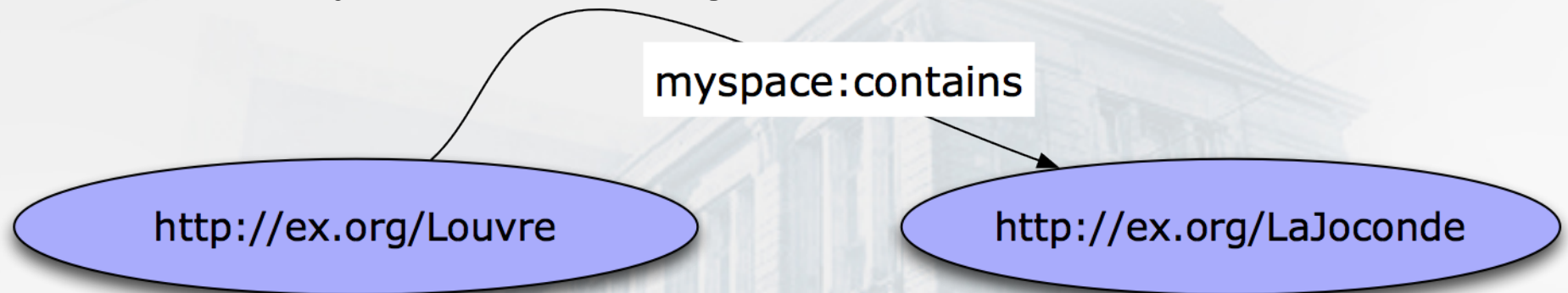
- We have HTTP URIs to identify resources and links between them – but we are missing a few things!
- What kinds of resources are 'Louvre.html' and 'LaJoconde.jpg'?
  - A machine cannot tell.
  - Humans can: we recognize implied context!
- How exactly do they relate to each other?
  - A machine cannot tell.
  - Humans can: again we recognize implied context!



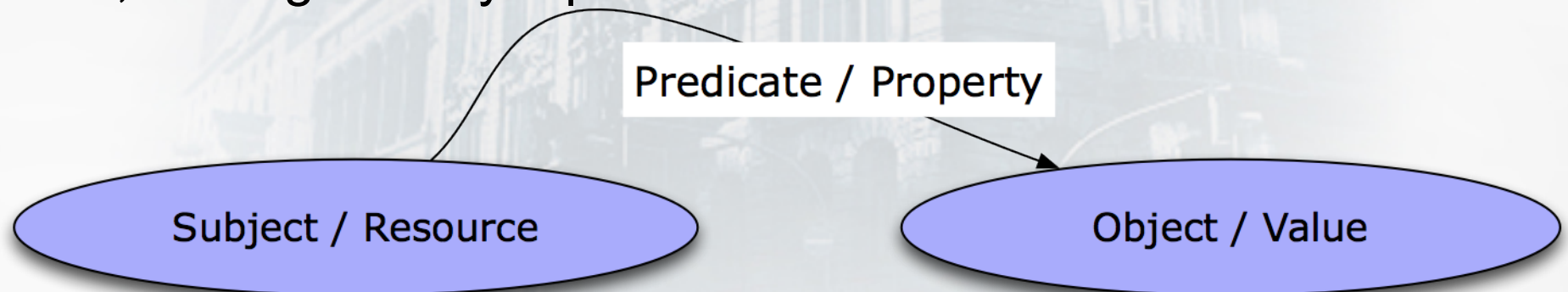
# Syntactically Extending the Document Web (1)



- We add a syntax for making statements on resources: RDF



- Or, more generally triples ...

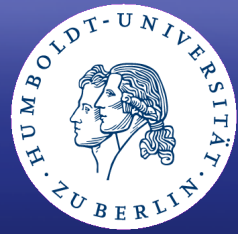


- ... where S and P are web resources (identified using URIs) and O is either a web resource or a literal

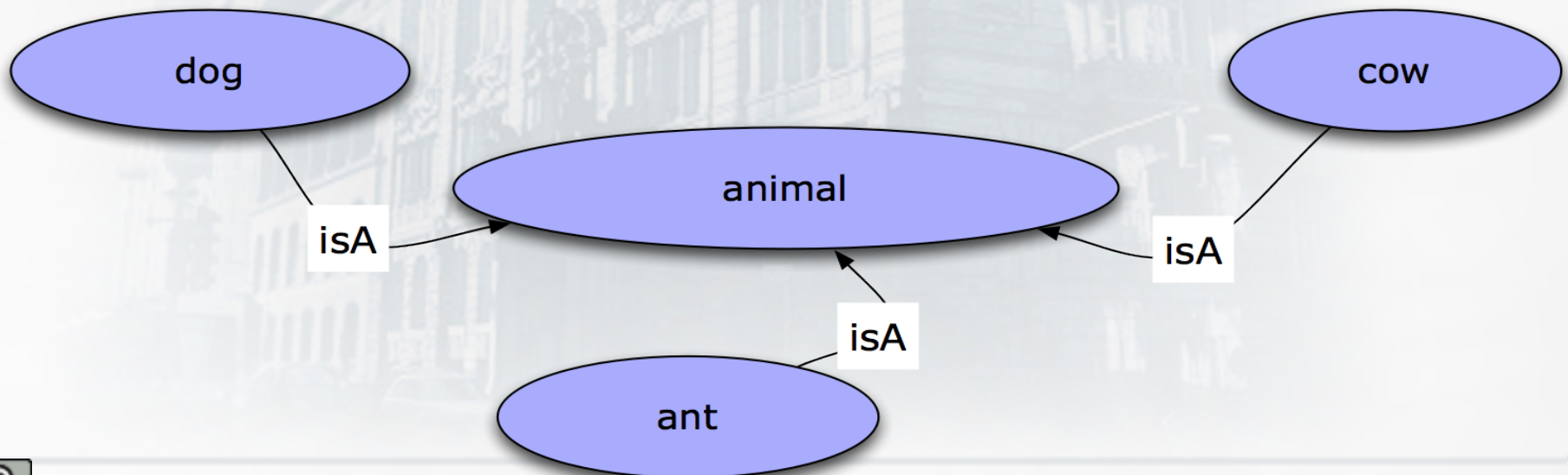




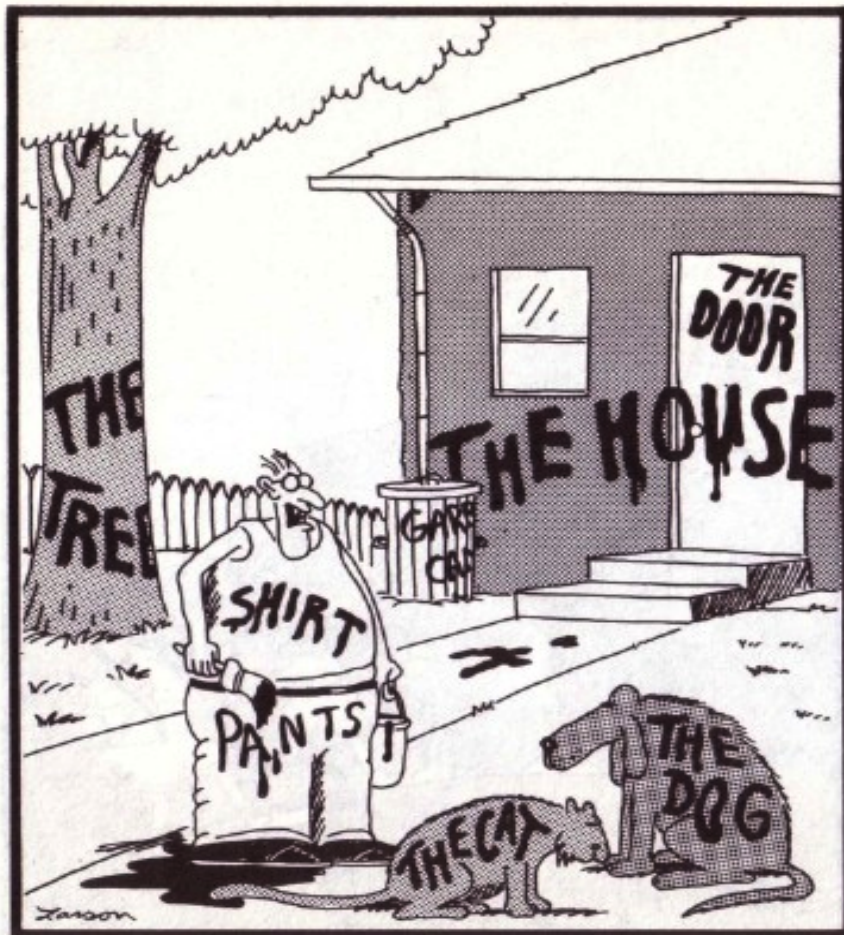
# Syntactically Extending the Document Web (2)



- We add a schema language (RDFS) with elements such as
  - classes,
  - hierarchies of classes and properties,
  - inheritance
  - support for basic inferencing.
- And thus are able to establish structures in triple aggregations resulting in lightweight domain ontologies:



# Extending the Web in Scope: The Web of Things ... (slightly Mistaken)



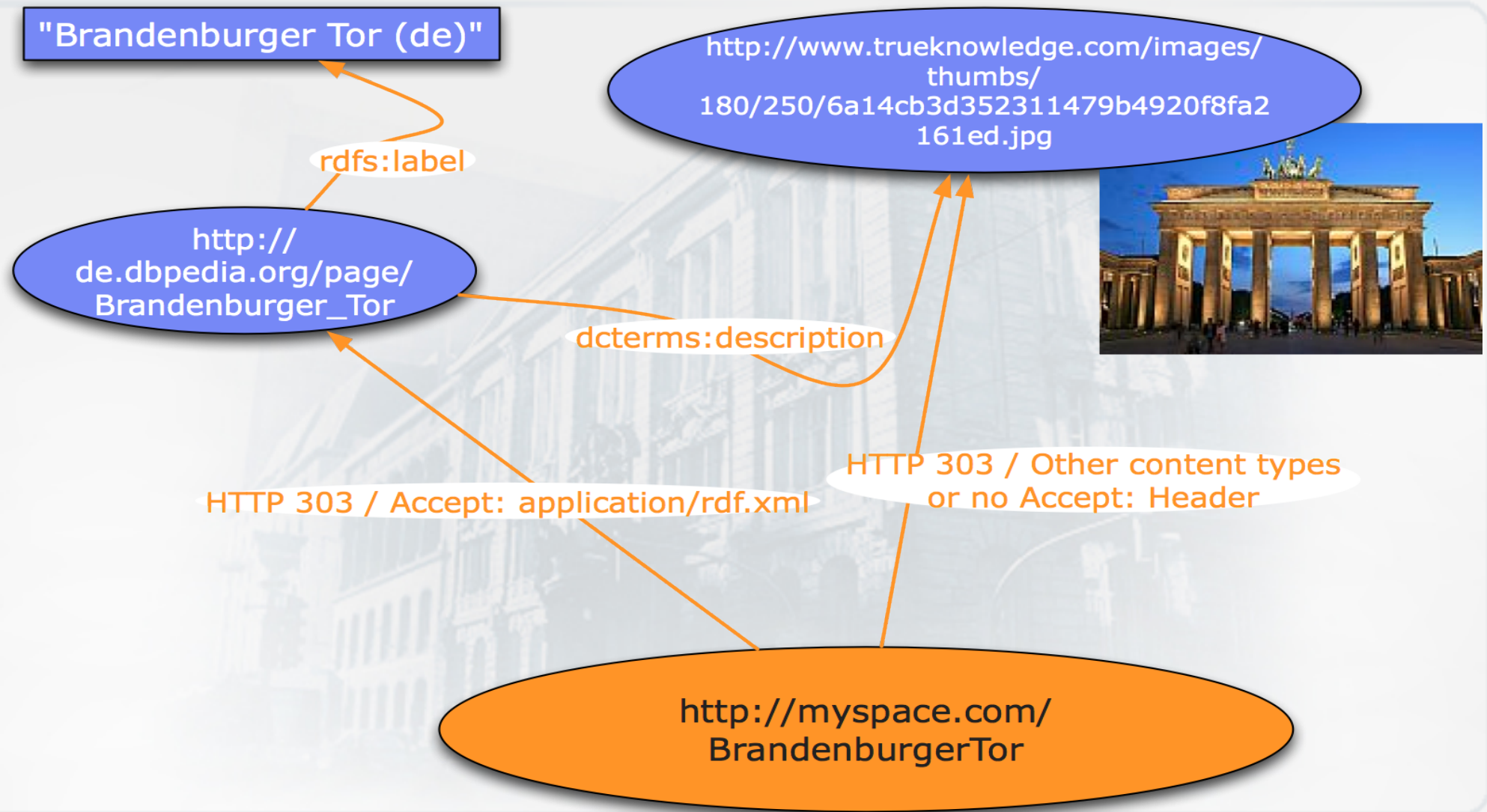
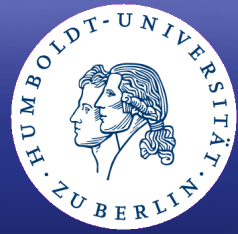
"Now! ... That should clear up  
a few things around here!"

What's **wrong**  
with this picture?

Taken from Ronald Carpentier's  
Blog at  
[http://carpentier.wordpress.com/  
2007/08/08/1-2-3/](http://carpentier.wordpress.com/2007/08/08/1-2-3/)



# ... and the Way we extend the Web in scope to make it a 'Web of Things'





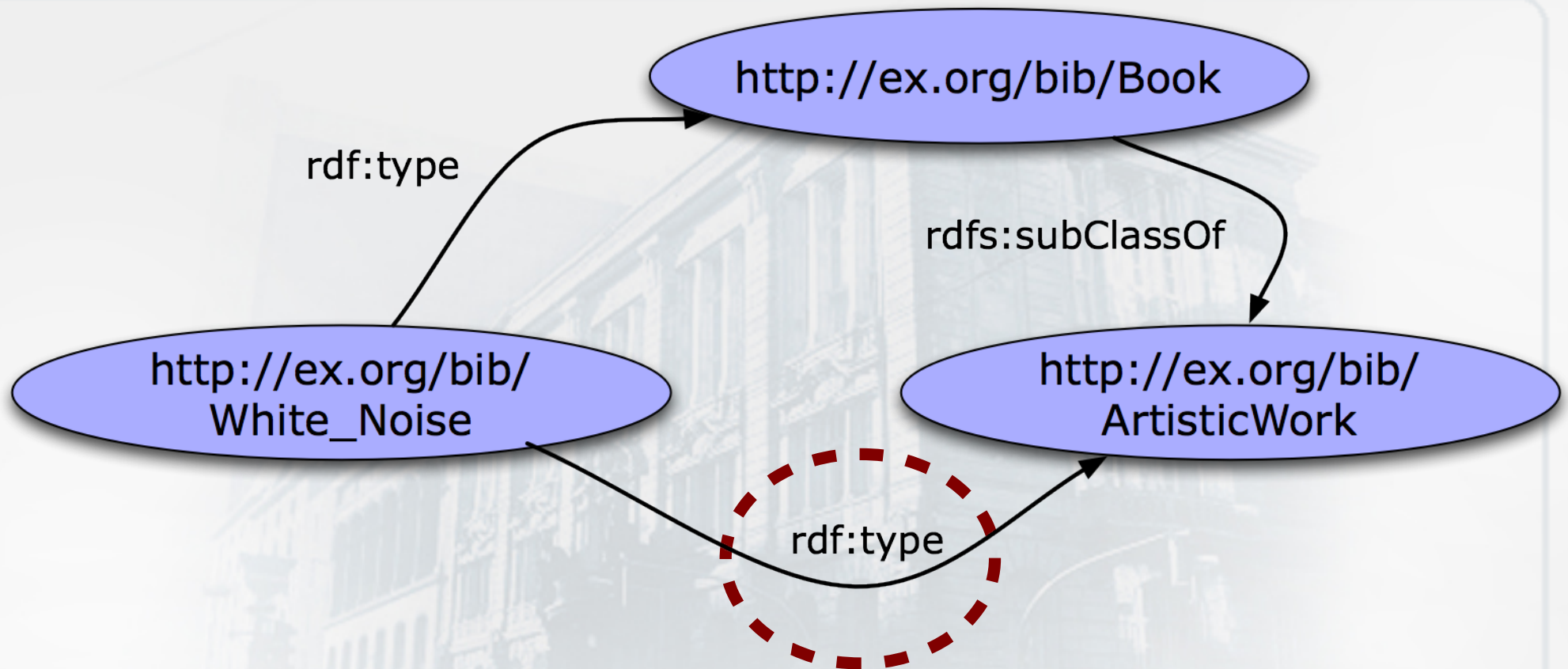
# Machines can reason on triple sets!



# Some reasoning preconditions ...



# ... and an automated inference!



There is quite some potential for generating scholarly heuristics here!





# Documents and Data in eScience and eScholarship

## Context is king!



# ... based on 'Documents' as Aggregations of RDF-Triples (1)



Assertion

NG\_000007.  
3:g.70628G>A

has  
frequency

0.25%

Condition

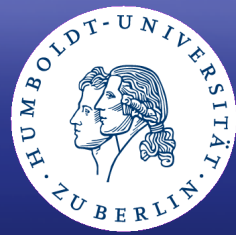
Sardinian

Provenance

Giardine et al



# 'Documents' as Aggregations of RDF-Triples (2)



```
<nanopublication id="0">
```

```
<assertion>
```

```
  <subject>NG_000007.3:g.70628G>A</subject>
```

```
  <predicate>has variant frequency</predicate>
```

```
  <object>0.25%</object>
```

```
</assertion>
```

```
<condition>Sardinian</condition>
```

```
<provenance>
```

```
  <dateofcreation>March 24, 2011</dateofcreation>
```

```
  <lastedit>March 24, 2011</lastedit>
```

```
  <evidenceType>empirical</evidenceType>
```

```
  <authorID>Giardine et. al.</authorID>
```

```
  <curatorID>unresolved</curatorID>
```

```
  <registrantID>Mons et. al.</registrantID>
```

```
  <PMID>6695908</PMID>
```

```
  <PMID>1428944</PMID>
```

```
  <PMID>1610915</PMID>
```

```
  <DOI>http://dx.doi.org/10.1038/ng.785</DOI>
```

```
  <linkout>http://globin.bx.psu.edu/cgi-bin/hbvar/query_vars3?
mode=output&display_format=page&i=239</linkout>
```

```
  <linkout>http://phencode.bx.psu.edu/cgi-bin/phencode/phencode?
build=hg18&id=HbVar.239</linkout>
```

```
</provenance>
```

```
<nanopublication id="0">
```





# The use of Inferences

Citation: van Haagen HHHBM, 't Hoen PAC, Botelho Bovo A, de Morrée A, van Mulligen EM, et al. (2009) Novel Protein-Protein Interactions Inferred from Literature Context. PLoS ONE 4(11): e7894. doi:10.1371/journal.pone.0007894 / Example provided by Jan Velterop



The screenshot shows the PLoS ONE article page. The title "Novel Protein-Protein Interactions Inferred from Literature Context" is highlighted with a red box. The authors listed are Herman H. H. B. M. van Haagen<sup>1\*</sup>, Peter A. C. 't Hoen<sup>1</sup>, Alessandro Botelho Bovo<sup>2</sup>, Antoine de Morrée<sup>1</sup>, Erik M. van Mulligen<sup>1</sup>, Christine Chichester<sup>1</sup>, Jan A. Kors<sup>1</sup>, Johan T. den Dunnen<sup>1</sup>, Gert-Jan B. van Ommen<sup>1</sup>, Silvére M. van der Maarel<sup>1</sup>, Vinícius Medina Kern<sup>2</sup>, Barend Mons<sup>1</sup>, and Martijn J. Schuemie<sup>1</sup>. The abstract states: "We have developed a method that predicts Protein-Protein Interactions (PPIs) based on the similarity of the context in which proteins appear in literature. This method outperforms previously developed PPI prediction algorithms that rely on the conjunction of two protein names in MEDLINE abstracts. We show". The page also includes a sidebar with metrics (Total Article Views: 673, Average Rating: 0), related content, and a share this article section.



# Semantic Publishing as Defined by Shotton

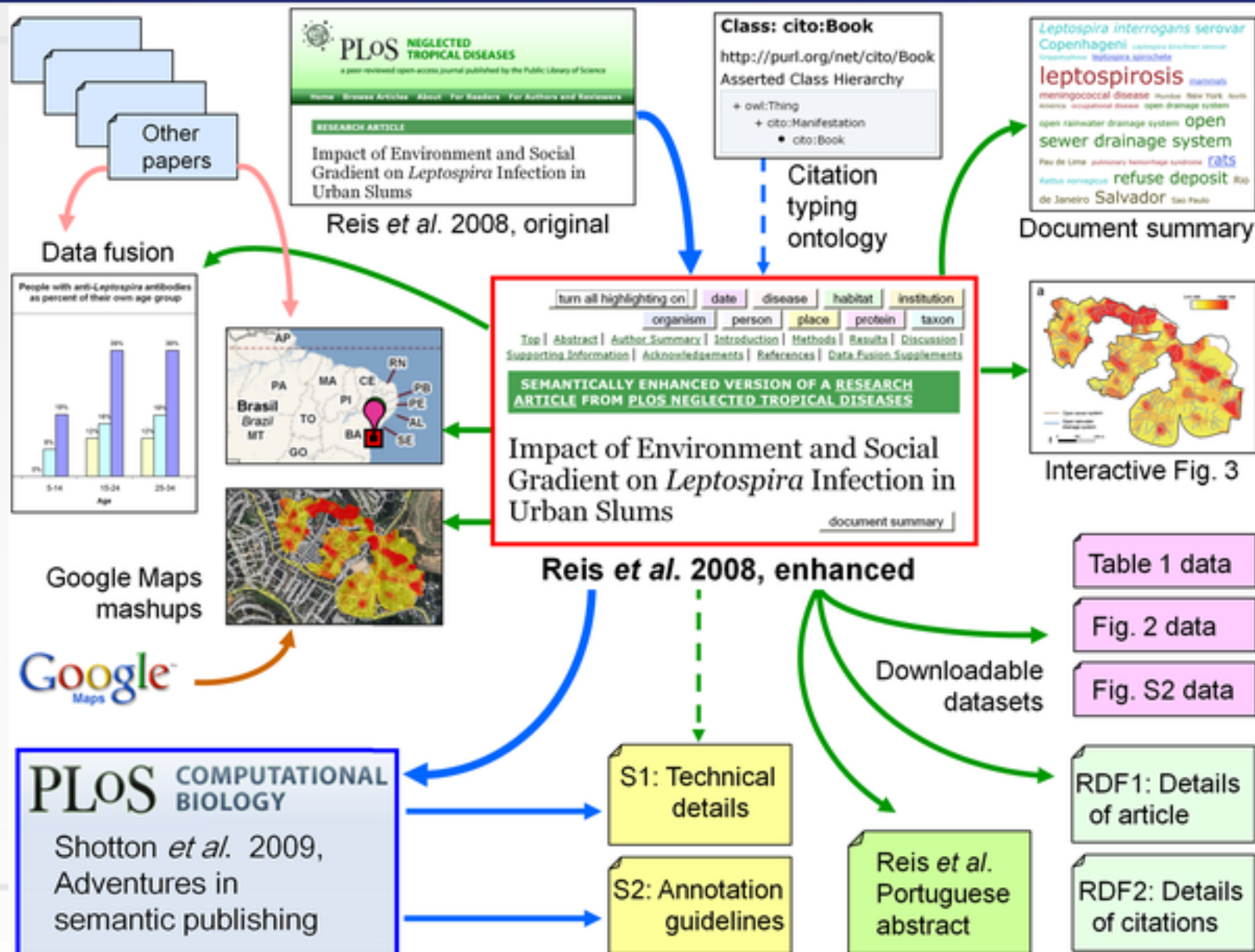


- Shotton et al. (2009b) define semantic publication to include anything that
  - **enhances the meaning** of a published journal article,
  - **facilitates** its automated **discovery**,
  - enables its **linking to semantically related articles**,
  - provides **access to data within the article** in actionable form, or
  - facilitates **integration of data between articles**.

Example of an enhanced article



# Behind the Screen





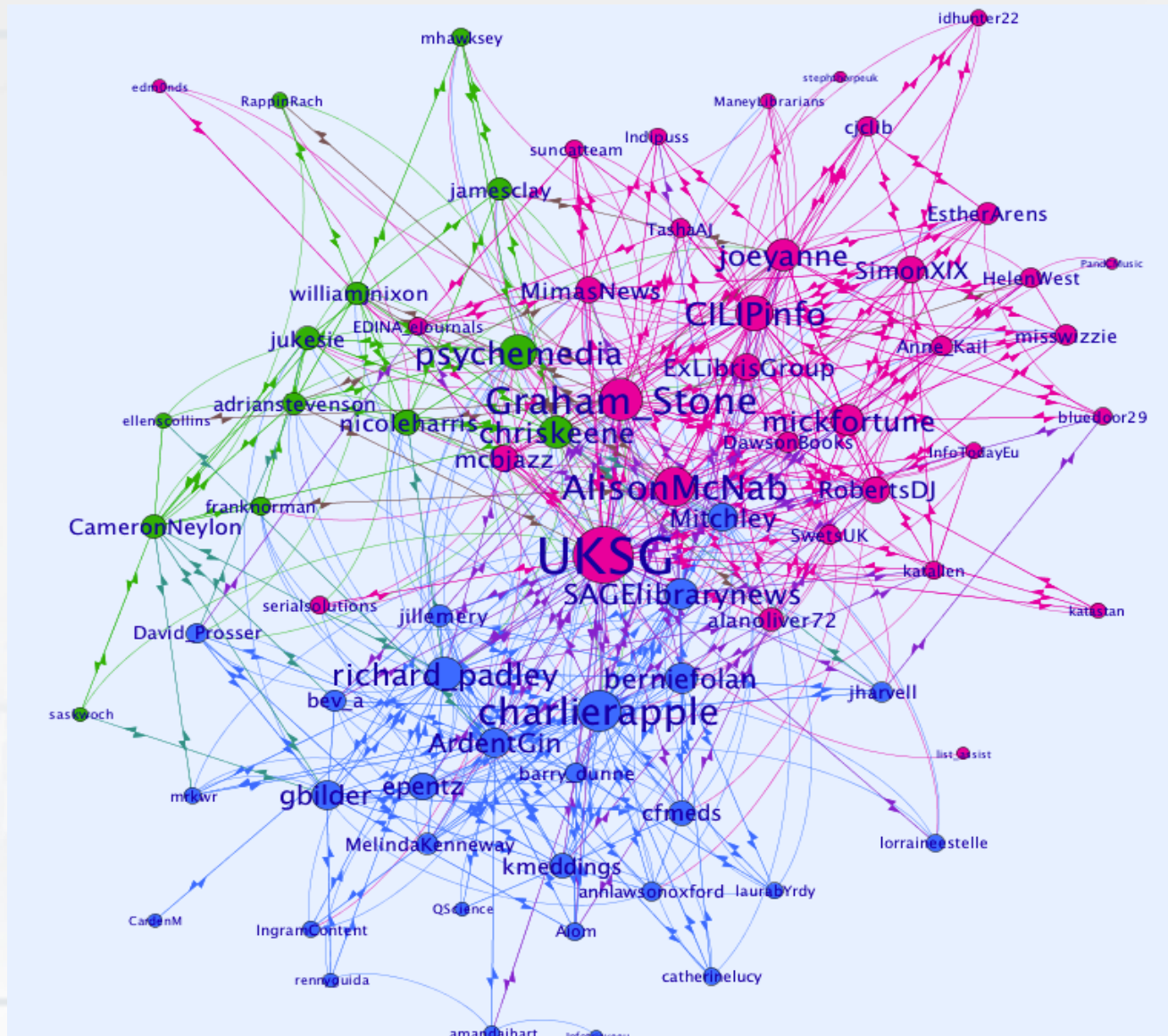
# Semantic Enrichment Tools

- Generic:
  - OpenCalais (<http://www.opencalais.com/> → Thomson Reuters)
  - Temis (<http://www.temis.com/>)
  - Collexis (<http://www.collexis.com/> → Elsevier)
- Specialised:
  - Bio Taxon Finder ([http://www.ubio.org/index.php?pagename=xml\\_services](http://www.ubio.org/index.php?pagename=xml_services))
  - ConceptWebAlliance (<http://conceptwiki.org>) (Biomedical, Jan Velterop)
- Shotton criticised by Roderic Page:  
<http://iphylo.blogspot.com/2009/04/semantic-publishing-towards-real.html>
  - “linking terms to HTML pages doesn't get us much further. Great for humans, not so good for computers.”
  - Too much focus on journal article format!
- → We need a little more! We need 'liquid documents'!!

# Data = Publication

- Distinction data vs. publication will get **increasingly obsolete** in semantic publishing environments ...
- ... at least in the STM sector.
- The move into semantic publication will be **much slower in the SSH** because of
  - fuzzy and unstable **terminology**
  - fuzzy **linking semantics** hard to formalise consistently
  - close relation between **complex document formats** and **scholarly discourse**
- Current examples are mostly from the medical and bio-medical area as a consequence.
- => Birte Christensen Dalsgard

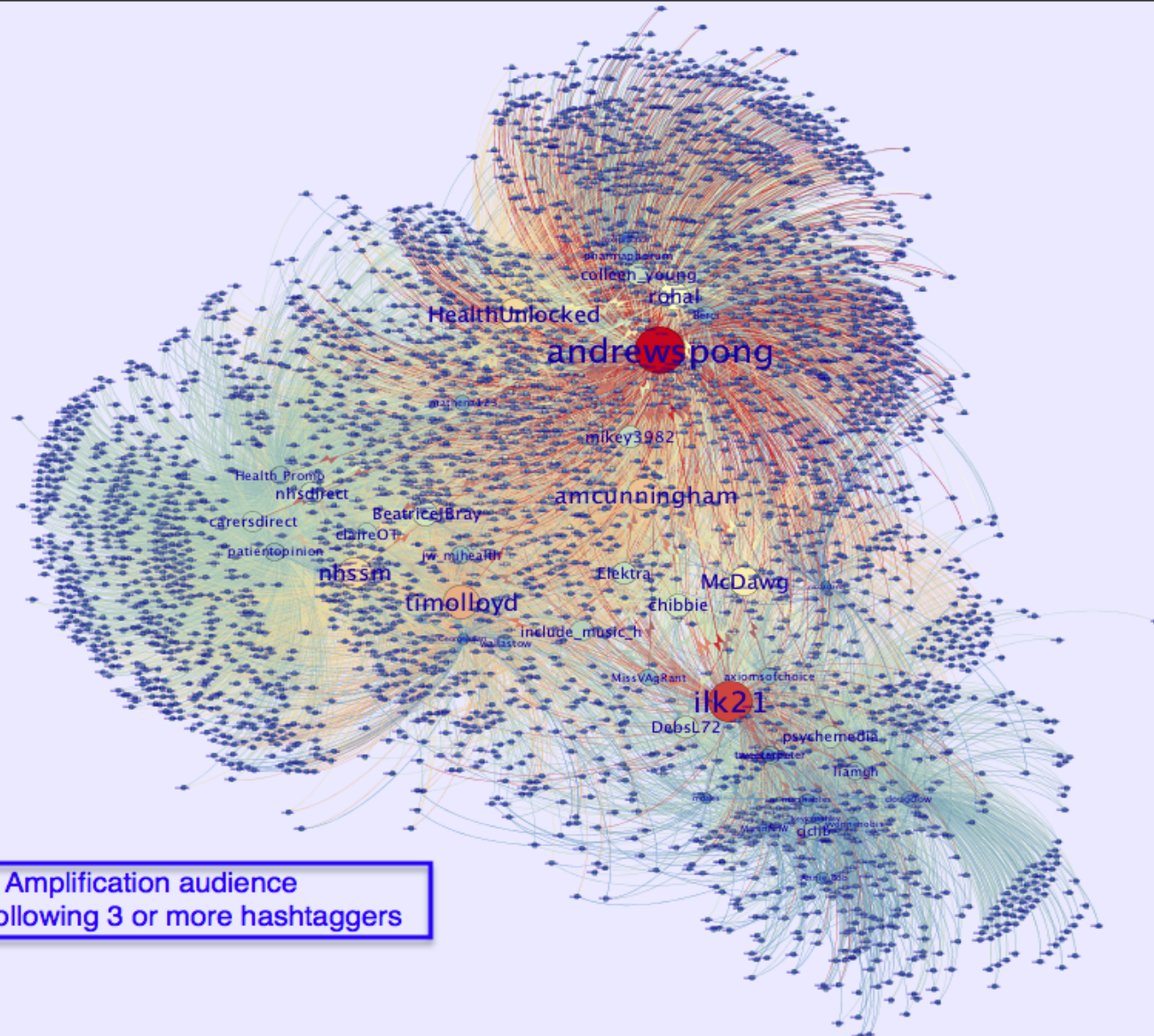
# ... visualise scholarly networks (1)



From Containers to Content to Context  
 Prof. Dr. Stefan Gradmann, Bielefeld Conference, April 24 2012



# ... visualise scholarly networks (2)



#iip11 Amplification audience  
Folk following 3 or more hashtags

# → Visualise Cultural Context

Mapping the Republic of Letters:

<https://republicofletters.stanford.edu/#maps>

Or again a Finnish example (Kulttuurisampo):

<http://www.kulttuurisampo.fi/kulsa/historiallisetKartat.shtml>

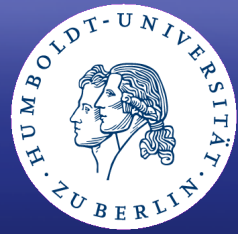


# **An Opportunity for Libraries ...** ... and what they need to do to be up to it





# “What do you do with a million books?” (G. Crane)



- Digitisation and (increasingly) semantic publishing result in
  - growing quantity
  - increased complexity
- Well beyond scholarly processing capacity (=reading faculty)
- Multiplication of collections or distributors is annoying → as few as possible. Ideally just one (?)
- Scientists and Scholars will badly need help in three areas:
  - **Semantic abstracting, named entity recognition** for “strategic reading” (Renear)
  - **Contextualisation** of information objects
  - Robust **reasoning and inferencing** yielding digital heuristics
- => ***Potential opportunities for libraries ...***



# Ceci n'est pas une bibliothèque

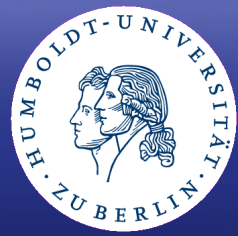


# Ceci n'est pas une bibliothèque

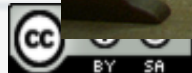




# Catalogue



The card catalog in the nave of Sterling Memorial Library at Yale University. Picture by Henry Trotter, 2005.





# Catalogue Entry: MARC Record



**LEADER:** 00495cam 2200169 a 4500  
001 94205672  
003 DLC  
005 20040108112243.0  
008 940907s1926 gw 000 1 ger  
010 \$a 94205672  
040 \$aDLC\$cDLC\$dDLC  
050 00 \$aPT2621.A26\$bS36 1926  
100 1 \$aKafka, Franz,\$d1883-1924  
245 14 \$aDas Schloss :\$bRoman /\$cFranz Kafka.  
260 \$aMünchen :\$bKurt Wolff Verlag,\$cc1926.  
300 \$a503 p. ;\$c20 cm.  
561 \$aSource: Bequest of Aaron Copland, Fall, 1989.\$5DLC

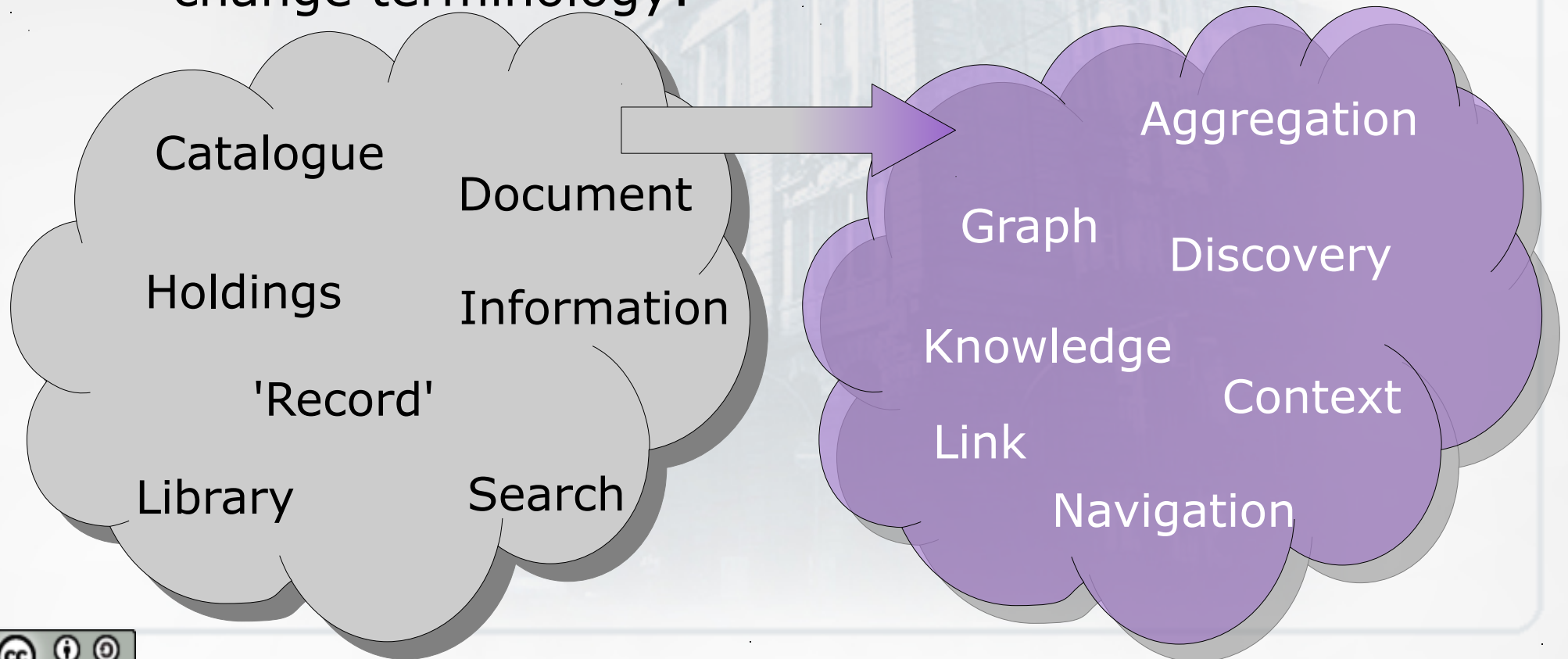


# 'Library Collections'

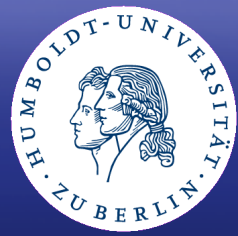


# Change Terminology!

- Libraries will serve research as part of the Linked Open Data web – or else risk becoming insignificant.
- For operating this change they definitely need to change terminology:



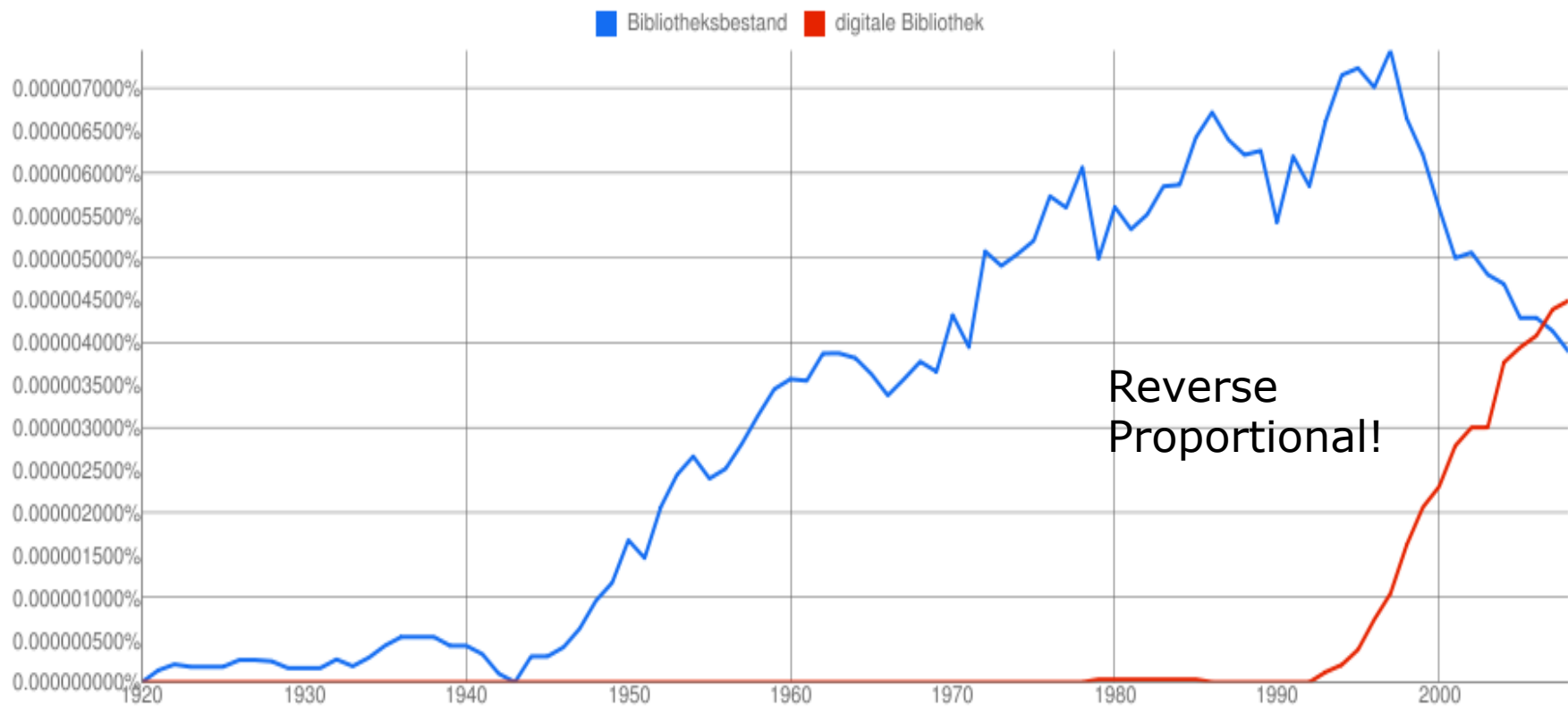
# From 'Catalogues' to 'Graphs': old terms – new terms (1)



Graph these **case-sensitive** comma-separated phrases:

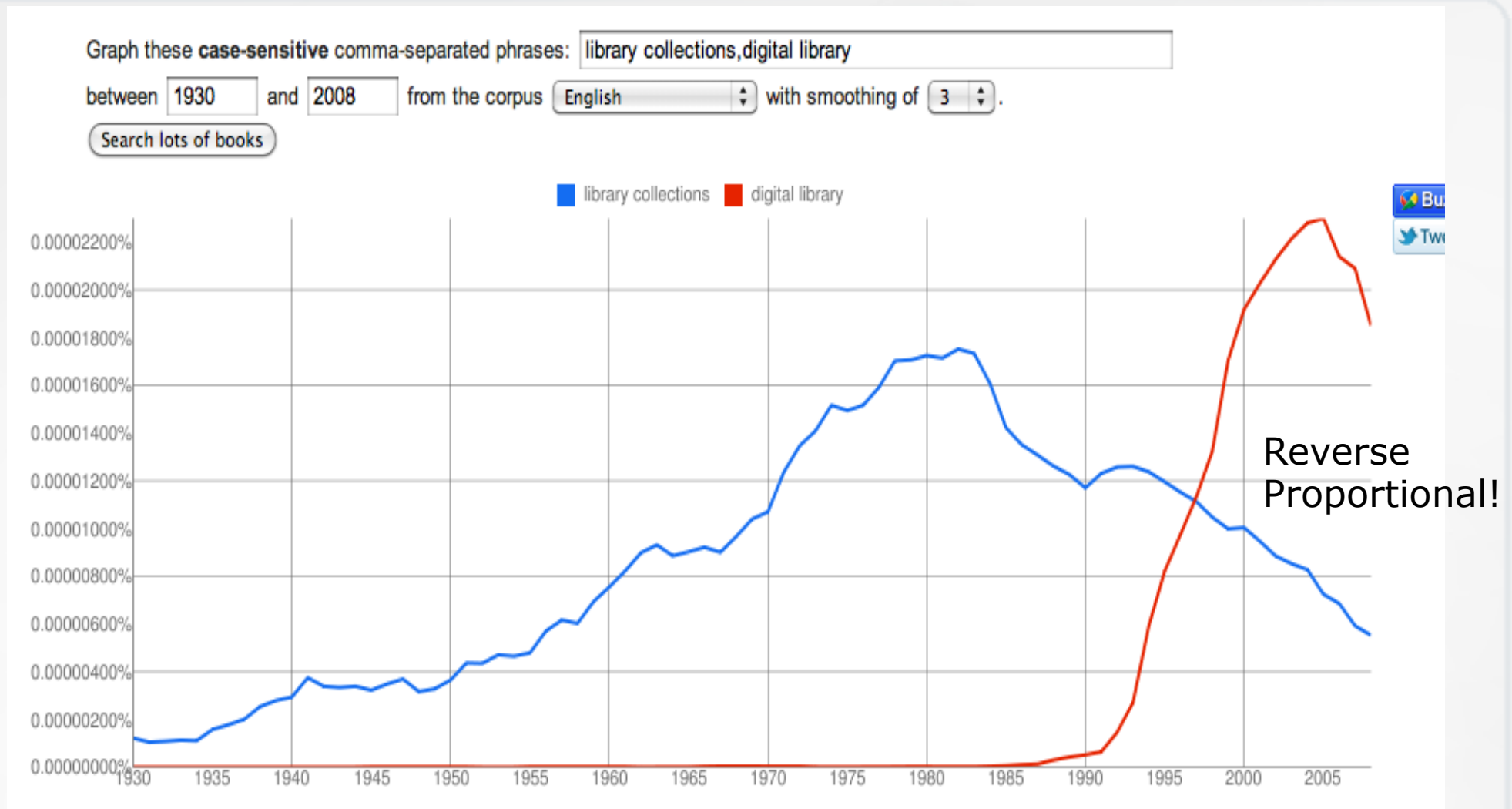
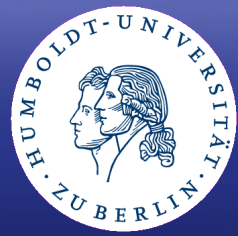
between  and  from the corpus  with smoothing of .

[Search lots of books](#)

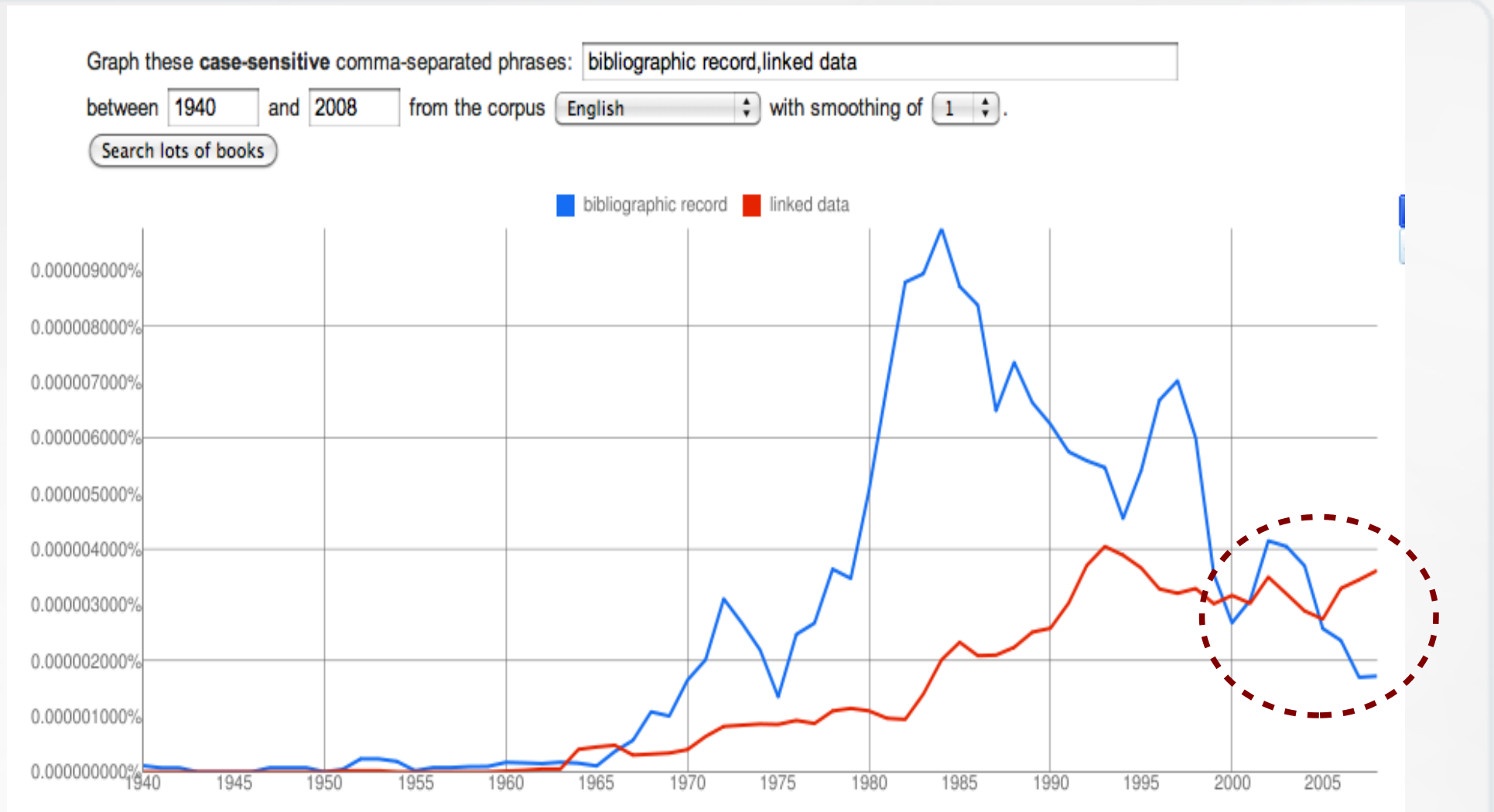
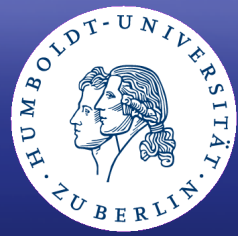




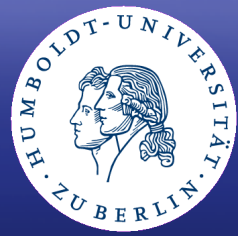
# From 'Catalogues' to 'Graphs': old terms – new terms (2)



# From 'Catalogues' to 'Graphs': old terms – new terms (3)



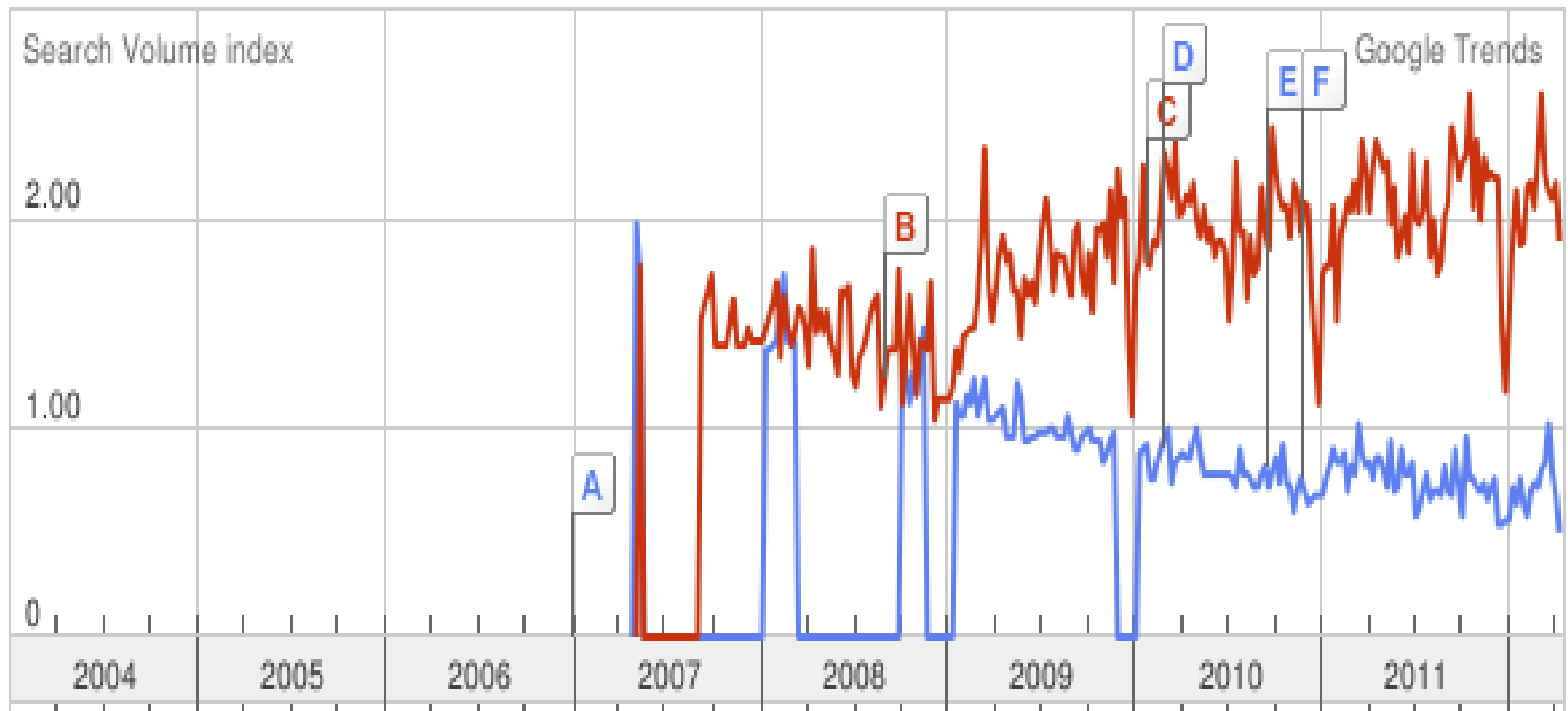
# From 'Catalogues' to 'Graphs': old terms – new terms (4)



cataloguing

0.38 linked data

1.00





# Sticking to empty metaphors ...

"What's in a name? That which we call a rose  
By any other name would smell as sweet."  
(Shakespeare, Romeo and Juliet (II, ii, 1-2))

- Why then do we stick to emptied metaphors?
- ... because they **constitute identity** (a very bad reason!)
- ... because they guarantee **institutional persistency** (a fallacy!)
- ... because we are afraid of substantial changes and believe in **things changing only once we use new terms** (dangerously childish!)
- ... or simply because **we do not have new terms** yet?

**Let us then start looking for them!**

# Suggested Reading

Gregory Crane (2006): What Do you Do with a Million Books? In: Dlib Magazine, Vol. 12, March. (<http://bit.ly/JhzF90>)

Gutenberg Paranthesis Research Group / University of Southern Denmark: Position Paper (<http://bit.ly/JjGKb6>)

David Parry: Burn the Boats/Books. Presentation to Digital Writing and Research Lab, Austin. (<http://bit.ly/JYLIJv>)

David Shotton (2009a): Semantic Publishing. The coming revolution in scientific journal publishing. Learned Publishing Volume 22, No 2, 85–94, April 2009; doi:10.1087/2009202

David Shotton et al. (2009b): Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article (<http://bit.ly/lgT5Km>)

Barend Mons, Jan Velterop: Nano-Publication in the e-science era (<http://bit.ly/lISMgt>)

Alan Renear, Carol Palmer (2009): Strategic Reading, Ontologies and the Future of scientific Publishing. In: Science, August 2009, p. 828 – 832.

**Thank you for your patience and attention**

